

Asymmetries of Men and Women in Selecting Partner

Haluk O. Bingol and Omer Basar

Department of Computer Engineering
Bogazici University, Istanbul

bingol@boun.edu.tr

November 7, 2012

Abstract

This paper investigates human dynamics in a large online dating site with 3,000 new users daily who stay in the system for 3 months on the average. The daily activity is also quite large such as 500,000 message transactions, 5,000 photo uploads, and 20,000 votes.

The data investigated has 276,210 male and 483,963 female users. Based on the activity that they made, there are clear distinctions between men and women in their pattern of behavior. Men prefer lower, women prefer higher qualifications in their partner.

1 Motivation

Until very recently we all live in a physical world only. With the rapid penetration of Internet to our daily lives starting in mid 1990's, we developed a new concept of virtual world. Now we have virtual friends that we exchange emails or tweets. We used to live in our social network, which was within our physical proximity. Now we talk about friends that we have not seen the face but know all of her ideas in almost all issues since we read his communications in the email list, or in her blog, or by means of her tweets.

Thanks to mobile Internet and new mobile devices such as smart phones and tablets we now have the ability to be connected to this virtual world 7-day-24-hour basis.

This is a new phenomenon that has never happened before. The impact of information and communication technologies (ICT) changed the way we live our lives [1]. The usual patterns of conducting our lives are in transition to adopt to the new

world of ICT. On the other hand, it provides invaluable datasets that have never been possible before [2, 3, 4, 5, 6, 7, 8, 9, 10]. Dating is not that can be kept away from this trend of virtualization [5].

Men and women are different in many ways [11, 12, 13, 14]. First of all they are different genetically due to X and Y chromosomes. One theory to explain genetic difference is the hunter-gatherer theory of Silverman which proposes a division of labor between genders, that is, men are hunters and women are foragers [11]. It has been shown that the brain of different genders works differently. For example, an electrophysiological study shows that men and women use different strategies, hence, different parts of their brain while judging beauty [13]. Men use mainly the right hemisphere for coordinate-based strategies. On the other hand women use both hemispheres for categorical strategies. They behave differently in many occasions. For example, men tend to have more sexual partners than women [12].

There are differences in gender but some of these differences may be due to social reasons such as historical or cultural. In late 1800, women are thought to be inferior to men in analytic abilities. As women get more share of everyday life, this view is changing. There is a clear trend of improvement of girls in test scores, such as SAT or PISA, as they take more science courses in the education system [14]. Once male dominated STEM fields, that is, the fields of science, technology, engineering, and mathematics, is becoming more even in genders.

Data for gender difference studies come from a number of resources. A survey based data collection is a method that is used frequently. As in any survey, the answers may be biased. Another data collection is

large-scale activities such as SAT tests. Recently the trace that we all leave on Internet becomes a source of data. In this paper, we investigate the data of a relatively large online dating system with focusing on the difference between men and women.

2 Online Dating Systems

A typical online dating system enables its user to find partner that best matches one’s desires. Each user defines his *user profile* in which he describes himself in terms of answering a set of questions such as gender, age, education, and income level. The questions are design in such a way that the answers are usually in the form of multiple choice rather than free format entries. User is asked to select one choice out of many.

A user searching for a partner defines what properties that he wants in his partner. This profile is called the *desired profile*. The same set of questions for profiling is used for this. The only difference is that one may select many out of multiple choices. For example, in his user profile he sets his education level as “university” while he can select “high school”, “masters” as well as “university” in the desired profile. Then all the users with profiles that match this are potential candidates.

3 The Dataset

The data that we investigate is based on an online dating system that is well established and one of the largest in Turkey.

There are 4,500,000 users registered. One should be careful about this big number. It is known that a user uses the system for three months on the average, than leaves it. Later, he may come back with his old user account. It is also possible that he creates a brand new account. There is no way for us to identify this new account to the old one. Therefore, this big number should not be understood as that many individual users. On the other hand, the system is used quite extensively. More than 3,000 new users registered daily. More than 50,000 users log on to the system daily.

The data investigated is huge. Considering the daily activity given in Table 1 and this activity is

Table 1: Daily activity numbers

Activity	Number
Messages sent	500,000
Gifts sent	1,000
Winks	10,000
Favorites	10,000
Votes	20,000
Photos uploaded	5,000
New users	3,000
Logins	50,000

accumulated through the years, the data size is in the order of couple of 100 GB.

The possible activities that users can make is also listed in Table 1. User can *wink* to a user to check if there is a potential interest. User can sent a *message* that is one-step higher involvement than a wink. It is possible that a *virtual gift*, images of flowers, can be sent. The difference of gifts from other actions is that the gifts are individually purchased while other actions are included in the membership fee. Therefore gift some how established a special value in the user community. A user can put another user in his *favorites* list. He can also *vote* to mean that he would go for her.

We investigate the data of $\mathcal{M} = 276,210$ male and $\mathcal{F} = 483,963$ female users. It may be surprising to see that there are more female in the data set, which calls for explanation. Due to the business model of the site, the system is free of charge for female users. On the other hand, there are two types of male users. The *free-of-charge male users* have restricted access to the facilities of the site. In order to get full access to the facilities of the site, male has to be a *paid member*. Our investigations require full access to facilities for both male and female users. Hence, we investigate all female users and only male users with paid membership. Since the free male users are not included, the number of male users is much less than that of female users.

As the investigation goes, some of the users do not satisfy the question at hand. Then the size of the sample space, that we get the statistics, reduces. For example, when we investigate the behavior of user that already sent a virtual gift twice, the size of the sample would be less than that of the sample who

both sent it twice and having an educational level of high school or higher. Since the number of male and female users is well above a few thousands even for the most restricted queries that we use, statistical values expected to be well founded. Since the number of female and male considered for a particular question changes by the question, we provide the number of male and female participant for that particular case by \mathcal{M} and \mathcal{F} values, respectively.

Clearly, privacy is the most important issue for such investigation. No individual personal information is used in this study. Data is randomized with MD5 algorithm for privacy concerns. No data left the company. All the data processing is done at the site of the company. Only statistical aggregated data is investigated in this study.

4 Observations

4.1 Profile

The mandatory fields of profile are birthday, marital status, town that you live in, your gender, and gender that you are looking for. Among the mandatory fields, the age has special importance since users under 18 are not allowed to use such systems by law.

Other properties of the profile are optional. User may or may not provide answers to them. A rich set of optional questions help user to define themselves. Among those are height, weight, eye color, hair color, body type, education level, occupation, position at work, salary, foreign languages, with whom he is living, parenthood status, child desire, smoking, alcohol status, as well as, the relation type that is sought for.

One needs to be careful that the profile is based on user's claims, that is, it may or may not be the real value of the user. On the other hand, cheating on the properties too far would be not a good strategy since when the time comes to meet face to face unfaithful declaration, such as declared as slim while being obese, would be obstacle to further the relationship. So we assume that user are closed to what they claim to be.

Observation 1. Although the system enables to select the exact height such as 178 cm, it is observed that users round up their height to the numbers multiple of 5, such as 170, 175, and 180. In Fig. 1 we see jumps at the multiples of 5. Clearly there is a

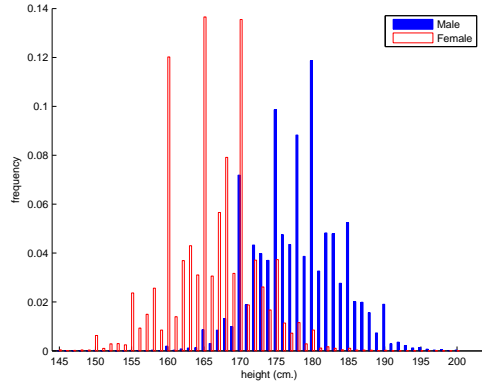


Figure 1: Users round up their height to the numbers multiple of 5. ($\mathcal{M} = 276,210$, $\mathcal{F} = 483,963$.)

Table 2: Percentages of answers by gender ($\mathcal{M} = 276,210$, $\mathcal{F} = 483,963$)

Property	Male	Female
Age	100	100
Education	99	97
Body type	99	97
Height	98	88
Salary	65	35

manipulation of information by the user. One may interpret this, as user prefers to use next level that he thinks better for him for the purpose. This is the case for both genders.

Observation 2. The first behavioral difference of the two genders that we observed is the percentage of the population that answers the profile questions. Although they are not required, majority of the users prefer to declare these fields. There is a clear difference in gender as seen in Table 2. Male does not hesitate to share information where as female is more concerned. For age, education and body type they behave almost the same but for height the difference starts. For salary, the percentage of male is almost twice of that of female.

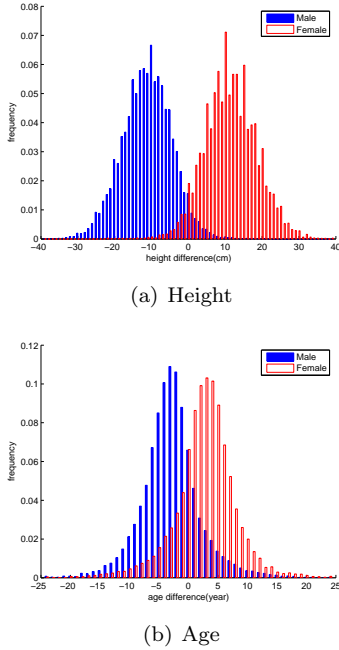


Figure 2: Distribution of difference in height and age. ($\mathcal{M} = 29,274$, $\mathcal{F} = 14,981$.)

4.2 Distribution of Property Difference

It is generally believed that people interact with people who have similar characteristics [15, 10]. In the context of this investigation, that means that someone, who claims to be 170 cm tall, is expected to become friend with the similar height.

The data enables us to quantify this. We restrict the data to the ones that *touch* to each other, that is, one sends a virtual gift and the other responds to this gift. Therefore both parties accepted the interaction. This restriction reduces the data set to $\mathcal{F} = 14,981$ female and $\mathcal{M} = 29,274$ male.

Consider an individual. The individual touches many people. The average height of all the people that the individual touches is calculated. The difference of the average and the height of the individual are obtained. If the individual has no tendency for taller or shorter friend, this difference is expected to be around zero. If it is positive than the individual prefers friends that are taller. In order to see if there

is a difference due to gender, individuals are grouped by gender. Then the individual tendencies are averaged by each gender group.

Fig. 2(a) provides tendency difference between genders. The difference distribution is bell shape as expected. The interesting observation is that the distributions are shifted with respect each other. The mean of the distribution is around -10 cm for men, whereas the mean for women is around $+10$ cm. So the conclusion is that men prefers shorter women. On the other hand women prefers taller men. It seems the ideal height difference between a man-women couple is expected to be around 10 cm.

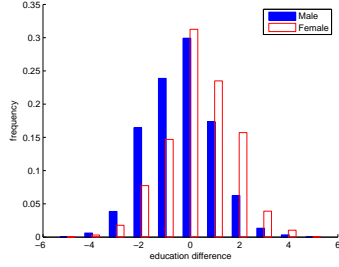
The distributions of men and women are similar in shape. They seems to be symmetric around the mean. Although both extends to -20 to $+20$ around the mean, the frequency drops sharply as one goes away from the mean. For example, men touching women taller than themselves are less than 5 %. Similarly, women preferring men shorter than themselves is also less than 5 %.

Similar approach is taken for other properties. Fig. 2(b), Fig. 3(a), Fig. 3(b), Fig. 3(c), and Fig. 3(d) provide gender differences in age, education, salary, body mass index (BMI), and body type, respectively.

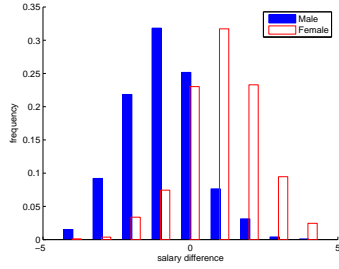
Observation 3. In all the distributions, one observes a similar shift due to gender. Men prefers negative, women prefers positive differences.

One should note that due to the design of the system, properties have two different natures. For some fields such as age and height, user is allowed to enter the exact value. For example, someone of height 182 cm, can enter exactly that value. For some other fields, user selects a range. Education, salary, body mass index (BMI), and body type are such fields. For example for the salary field, user is allow to select one out of five salary bins, namely, bin 1: $x < 500$, bin 2: $500 < x < 1000$, bin 3: $1000 < x < 2000$, bin 4: $2000 < x < 3000$, and bin 5: $3000 < x$. So difference in the salary actually means the difference between the bin numbers not the actual salaries. Fortunately, the mapping from the actual value to the bin number is order-preserving mapping. Hence, the difference still has a meaning.

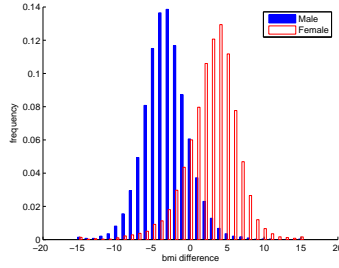
Observation 4. Note a very interesting point that both genders prefer the same amount of difference in all properties. Consider a property such as height. Suppose one gender prefers 10 cm of height differ-



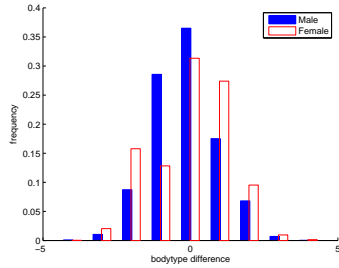
(a) Education



(b) Salary



(c) BMI



(d) Body Type

Figure 3: Distribution of difference in educational level, salary, BMI and body type. ($\mathcal{M} = 29,274$, $\mathcal{F} = 14,981$.)

Table 3: Predefine wink messages.

No	Message
1	I loved your photo. You are cool!
2	I liked your photo. You smile nice!
3	I liked your profile. Why don't you have a photo?
4	I liked your profile. Tell me about yourself.
5	We are suitable for each other. Look at my profile.
6	Everything starts with a hello. Hello!
7	If you're looking for a funny friend, I am here.
8	I don't want to add a message.

ence while the other 15 cm. Then, very few of us would be happy. This is, perhaps, due to an evolutionary adaptation which calls for an evolutionary explanation similar to that of ref [11].

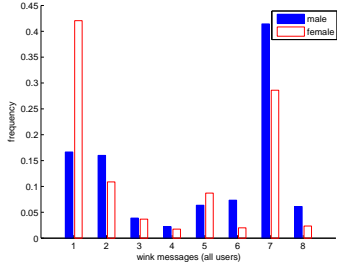
4.3 Distribution of Wink Messages

The system provides a set of predefined messages, called *wink*, that a user can send to. The predefined wink messages are given in Table 3. $\mathcal{M} = 216,891$ male and $\mathcal{F} = 241,748$ female users are included in this study.

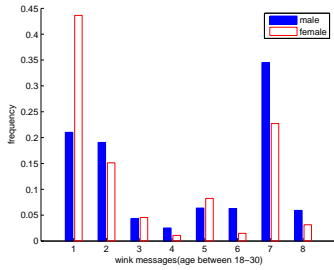
Observation 5. Usage of wink itself has asymmetry. Male users use winking 10 times more than female users. This can be interpreted as male takes the first move.

Selection of the wink message is also asymmetric as seen in Fig. 4. Fig. 4(a) provides message selection for all ages combined. The message “7: *If you're looking for a funny friend, I am here.*” is the most preferred message by male users. Interestingly, in almost one of every two winks by men, this message is selected. For female users this message is the second most preferred message. The most frequent message of female users is “1: *I loved your photo. You are cool!*” with more than 40% usage.

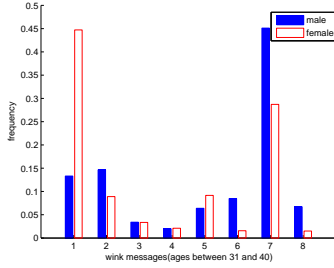
It is observed that selection changes by age in both gender. Fig. 4(b), Fig. 4(c), and Fig. 4(d) provide change in selection for different age groups of 18–30, 31–40, and 41–50, respectively. As men gets older,



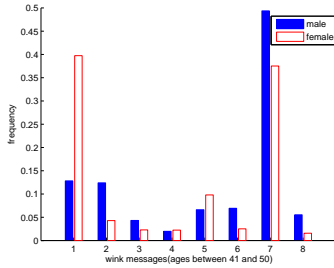
(a) Education



(b) Salary



(c) BMI



(d) Body Type

Figure 4: Distribution of wink message usage. ($\mathcal{M} = 216, 891$, $\mathcal{F} = 241, 748$.)

they tend to use message “7: *If you’re looking for a funny friend, I am here.*” more. Its frequency goes up from 35% to almost 50%. There is no clear second message. The second and the third frequently used messages, namely, “1: *I loved your photo. You are cool!*” and “2: *I liked your photo. You smile nice!*” are used almost at the same frequencies.

For the female users, the most frequently used message is “1: *I loved your photo. You are cool!*”. As women get older, its frequency drops nearly from 45% to 40%. In contrast to males, there is a clear second most frequently used message for female users. It is “7: *If you’re looking for a funny friend, I am here.*”, which was the favorite of males. Its usage increases approximately from 23% to 37% as women gets older.

For the mature age group of 41 – 50, the first two messages hold more than 75% usage for females where as the same number is much lower, around 63% for males. At all age groups, the coverage of the first three messages is much higher for females than males, that is, above 80% and 75%, respectively.

Observation 6. This can be summarized as, visual appearance is more important than profile. “Fun” is always important and becomes increasingly more important for both genders as they get older. Women decisively prefers to have “cool photo” as a initial message for friendship attempt as oppose to men prefers equally “cool photo” and “nice smile” messages.

5 Discussion

In many respects men and women behave differently [13]. In the virtual world of online dating is another manifestation of this difference. It is found that men take the first move in partner selection. Then, women have the right to accept or reject. Men are more open to share their private information than women. While male prefers women with lower qualifications such as income level, women do the opposite. Women prefer men of higher qualification such as higher education level. Both genders like funny partner yet the level of impact differs by gender and also by age with in the same gender.

Acknowledgments. This work was partially supported by Bogazici University Research Fund, BAP-2008-08A105, by the Turkish State Planning Orga-

nization (DPT) TAM Project, 2007K120610, and by COST action MP0801.

References

- [1] Albert Ali Salah, Bruno Lepri, Fabio Pianesi, and Alex Sandy Pentland. Human Behavior Understanding for Inducing Behavioral Change: Application Perspectives. In *HBU 2011, LNCS 7065*, pages 1–15, 2011.
- [2] Peter Sheridan Dodds, Roby Muhamad, and Duncan J Watts. An experimental study of search in global social networks. *Science (New York, N.Y.)*, 301(5634):827–9, August 2003.
- [3] Jure Leskovec, Ajit Singh, and Jon Kleinberg. Patterns of Influence in a Recommendation Network. *Advances in Knowledge Discovery and Data Mining, LNCS*, 3918:380–389, 2006.
- [4] J-P Onnela, J Saramäki, J Hyvönen, G Szabó, D Lazer, K Kaski, J Kertész, and A-L Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7332–6, May 2007.
- [5] J. Barraket and M. S. Henry-Waring. Getting it on(line): Sociological perspectives on e-dating. *Journal of Sociology*, 44(2):149–165, June 2008.
- [6] Diego Rybski, Sergey V Buldyrev, Shlomo Havlin, Fredrik Liljeros, and Hernán a Makse. Scaling laws of human interaction activity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):12640–5, August 2009.
- [7] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–7, September 2010.
- [8] Luis E C Rocha, Fredrik Liljeros, and Petter Holme. Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):5706–11, March 2010.
- [9] M. Szell, R. Lambiotte, and S. Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31), July 2010.
- [10] Damon Centola. An experimental study of homophily in the adoption of health behavior. *Science (New York, N.Y.)*, 334(6060):1269–72, December 2011.
- [11] I Silverman and M Eals. Sex differences in spatial abilities: Evolutionary theory and data. In Jerome H Barkow, Leda Cosmides, and John Tooby, editors, *The Adapted Mind Evolutionary Psychology and the Generation of Culture*, volume The adapte, pages 533–549. Oxford University Press, 1992.
- [12] F Liljeros, C R Edling, L a Amaral, H E Stanley, and Y Aberg. The web of human sexual contacts. *Nature*, 411(6840):907–8, June 2001.
- [13] Camilo J Cela-Conde, Francisco J Ayala, Enric Munar, Fernando Maestú, Marcos Nadal, Miguel a Capó, David del Río, Juan J López-Ibor, Tomás Ortiz, Claudio Mirasso, and Gisèle Marty. Sex-related similarities and differences in the neural correlates of beauty. *Proceedings of the National Academy of Sciences of the United States of America*, 106(10):3847–52, March 2009.
- [14] Janet S Hyde and Janet E Mertz. Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences of the United States of America*, 106(22):8801–7, June 2009.
- [15] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444, August 2001.